

Partner Read Aloud: Data Checking at Its Fastest

Marlen Ibarra, Sarah Willick, Amber K. Stephens, Brianna R. Maxim, & Kimberly A. Barchard

University of Nevada, Las Vegas



Abstract

Data verification is an essential part of data management. Some data checking methods take longer than others. This study compared the speed of four data checking methods: visual checking, solo read-aloud, partner read-aloud, and double entry. A total of 154 undergraduates completed this study for course credit. They checked the data from 20 data sheets that had previously been entered into Excel. The Excel file contained errors, and participants' job was to find and correct those errors. We found double entry was significantly slower than all of the other methods and solo read aloud was slower than partner read aloud. Given that it takes longer, double entry may cost more than other techniques and may also be more physically and mentally taxing. However, previous studies have found that double entry is the most accurate. Whether double entry is worth the extra time will depend upon the context of a study.

Introduction

Detecting and correcting data entry errors is an essential part of data management

(Chen, Chen, Conway, Hellerstein, & Parikh, 2011). Data checking is used in medical, academic, and non-academic settings (Johnson, Temple, Pearson & Weber, 2009). If data checking takes a long time, the people doing the data checking may make more errors because they are bored or tired (Healy, Kole, Buck-Gengler, & Bourne, 2004) or their eyes get sore (Rechichi, De Moja, & Scullica, 1996). Thus, using fast data checking methods contributes to cost-effectiveness and timeliness (Norr, 2000).

Three studies have compared the speed of various data checking methods. They have found that double entry takes more time than read aloud (Barchard & Verenikina, 2013; Reynolds-Haertle & McBride, 1992) and visual checking (Barchard & Verenikina, 2013; Kawado, Hinotsu, Yamaguchi, Hashimoto, Matsuyama, & Ohashi, 2003; Reynolds-Haertle & McBride, 1992). However, no studies

ID: 739925	Sex: (M) F
Learning Style	Study Habits
1. 1 2 3 (4) 5	1. SD D N A (S)
2. 1 2 3 4 (5)	2. SD D (N) A SA
3. 1 2 3 4 (5)	3. SD D N A (S)
4. 1 2 (3) 4 5	4. SD D N A (S)
5. 1 2 3 4 (5)	5. SD D N (A) SA
6. 1 2 3 (4) 5	6. SD D N (A) SA
7. 1 2 3 4 (5)	7. SD (D) N A SA
8. 1 2 (3) 4 5	8. SD (D) N A SA
Spelling Test	Math Test
1. ACCOMMODATE	1. 156
2. AMATEUR	2. 235
3. CALENDAR	3. 485
4. CEMETERY	4. 493
5. CONSHENCE	5. 364
6. EMBARRASS	6. 327
7. EXHILARATE	7. 203
8. MAINTAINANCE	8. 347

have compared the speeds of solo read aloud and partner read aloud. Therefore, the purpose of this study is to compare the speed of double entry, visual checking, solo read-aloud, and partner read-aloud. Based upon previous research, we hypothesize that visual checking will be the fastest.

Method

Participants

A total of 154 undergraduate students (103 female, 49 male, 2 unidentified) participated in this study in return for course credit. Their ages ranged from 18 to 50 (mean 21.56, SD 6.42). The participants self-identified their ethnicities as follows: 10.4% African American, 20.1% Asian, 36.4% Caucasian, 26.0% Hispanic, 0.6% Native American, 1.9% Pacific Islander, and 3.9% Other.

Procedures

Before participants arrived, 25 data sheets (see Figure 1) were entered into Excel. However, when we entered the data, we deliberately introduced data entry errors. Participants' job was to locate and correct these errors.

Participants completed the study in a single 90-minute supervised session. Because the data checking would use the Excel program, participants were asked to watch a video tutorial on the basics of Excel. Next, participants were randomly assigned to one of four data checking methods (described below), and watched a second video that explained this method.

Method Cont.

During part 1 of the study, the participants checked the data for the five data sheets while

the administrator observed them and corrected any errors the participants made. During part 2, participants used this same data checking method to check the data for 20 additional data sheets, without assistance from the administrator. The website on which the study was conducted collected the start and end times of the second part of the study in order to calculate how long the data checking took.

Participants were randomly assigned to one of four data checking methods. In visual checking, the participant held the paper data sheet and visually compared it to the data on the screen. In solo read-aloud, the participant read the data sheet out loud while simultaneously checking the entry on the computer. In partner read-aloud, one person reads the information from the data sheet out loud and the other compares the spoken entry to the information on the screen. In this study, the administrator of the study read the paper data sheet and the participant verified the entry in Excel. In double entry, the participant entered the data into a blank sheet in the Excel file and the computer compared these new entries to the existing entries to determine if there were any mismatches between the entries. The computer also checked for any values that were outside the allowable range for those variables. Regardless of which method the participant used, they were asked to correct any data entry errors that they found.

Data Analysis

We compared how long it took participants to check the data using the four different data checking methods using a one-way ANOVA. We compared each pair of conditions using Tukey's HSD.

Results

There was a significant difference in the time it took people to complete data checking under the four data checking methods ($F(3, 150) = 32.58, p < .001$). Pairwise comparisons using Tukey's HSD revealed that double entry was significantly slower than all of the other methods, and that solo read aloud was slower than partner read aloud. See Table 1 and Figure 2.

Discussion

This study's primary focus was to determine which of the four data checking methods was the fastest. Contrary to our hypothesis, the fastest data checking method was partner read-aloud. Partner read aloud may have been faster than solo read aloud and visual checking because the participant did not have to read the paper data sheet – that was done by the administrator. This also helped participants focus their attention on the Excel sheet. Additionally, during partner read aloud, the administrators read the entries at a consistent tempo, whereas participants in the solo read aloud condition may have slowed down their reading rate when they encountered possible errors. Norr (2000) argues that the time spent performing data checking does not exclusively depend on the method but varies due to the participant's ability to check the data quickly. Because the researchers kept a consistent pace while reciting the entries, individual variability was removed, creating a faster average time than the other methods.

In past studies, double entry has been the most accurate (Barchard & Pace, 2008; Barchard & Verenikina, 2013; Reynolds-Haertle & McBride, 1992; Kawado et al., 2003); however, with an average time of 46.7 minutes in our study, it is the slowest of the data checking methods.

Whether accuracy should be considered more important than speed depends upon the context. For example, Barchard and Pace (2008) demonstrated that data entry errors can drastically alter the results of a study and concluded double entry was worth the extra time. On the

Discussion Cont.

other hand, Johnson et al (2011) found that data entry errors did not influence the results enough to warrant the extra time and energy with fisheries data. When deciding what data checking method to use, database managers should consider the frequency and type of errors that are likely to be made and the consequences of those errors.

Finally, our study used undergraduate students, who probably have little data checking experience. Future studies should examine participants who have previous data checking ex-

perience; experience may change the relative speed of the methods. This could alter evaluation of the speed-accuracy tradeoff.

References

Barchard, K. A., & Pace, L. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior, 27*, 1834–1839. doi:10.1016/j.chb.2011.04.004

Barchard, K. A., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. *Computers in Human Behavior, 29*(5), 1917–1922. doi:10.1016/j.chb.2013.02.021

Beatty, J. (1999). The powerchecker: A visual basic program for ensuring data integrity. *Behavior Research Methods, In-*

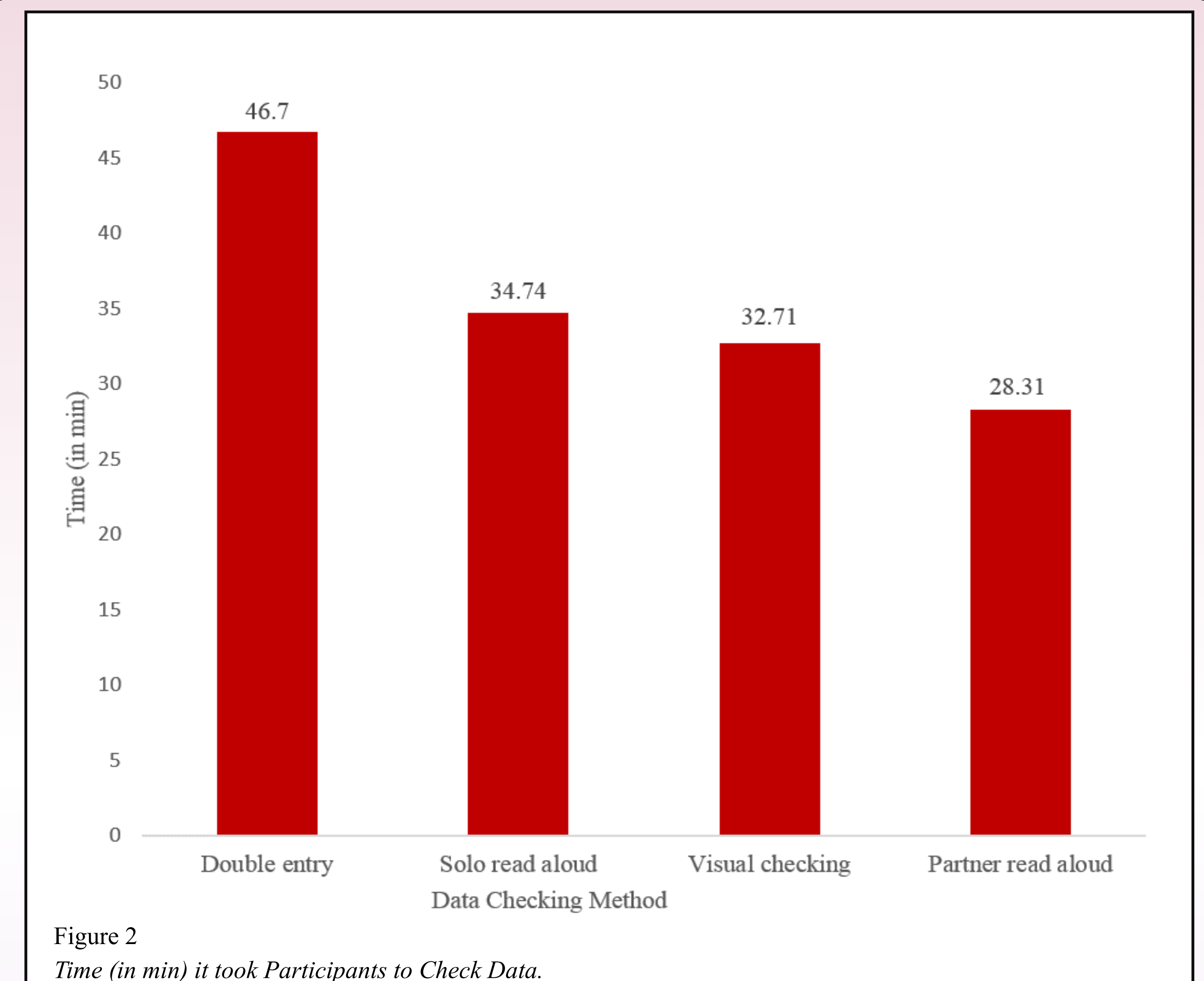


Figure 2
Time (in min) it took Participants to Check Data.

struments, & Computers, 737–740. doi:10.3758/BF03200757

Chen, K., Chen, H., Conway, N., Hellerstein, J. M., & Parikh, T. S. (2011). Usher: Improving data quality with dynamic forms. *Knowledge and Data Engineering, IEEE Transactions on, 23*, 1138–1153. doi:10.1109/TKDE.2011.31

Healy, A. F., Kole, J. A., Buck-Gengler, C., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10*(3), 188–199. doi:10.1037/1076-898X.10.3.188

Johnson, C. L., Temple, G. M., Pearsons, T. N., & Webster, T. D. (2009). An evaluation of data entry error and proofing methods for fisheries data. *Transactions of the American Fisheries Society, 138*, 593–601. doi:10.1577/T08-075.1

Kawado, M., Hinotsu, S., Yamaguchi, T., Hashimoto, S., Matsuyama, Y., & Ohashi, Y. (2003). A comparison of error detection rates between the reading aloud method and the double data entry method. *Controlled Clinical Trials, 24*, 560–569. doi:10.1016/S0197-2456(03)00089-8

King, D. W., & Lashley, R. (2000). A quantifiable alternative to double data entry. *Controlled Clinical Trials, 21*(2), 94–102. doi:10.1016/S0197-2456(00)00042-8

Norr, K.B. (2000). An investigation of the relationship between speed and accuracy utilizing a data entry task. *Dissertation Abstracts International: Secion B: The Science and Engineering, 60*, 3606.

Rechichi, C., De Mojà, C. A., & Scullica, L. (1996). Psychology of computer use: XXXVI. Visual discomfort and different types of work at videodisplay terminals. *Perceptual and Motor Skills, 82*, 935–938. doi:10.2466/pms.1996.82.3.935

Reynolds-Haertle, R. A., & McBride, R. (1992). Single vs. Double data entry in CAST. *Controlled Clinical Trials, 13*, 487–494. doi:10.1016/0197-2456(92)90205-E